

# Smart E-Healthcare System: A Web-Based Multi-Disease Prediction Platform Using Multi-Label Machine Learning with Symptom-Based Diagnosis and Integrated Clinical Decision Support

Vattikoti Sai, Fathima Zahera, Manikanta, Shiva, Rakesh, Bhuvana Sri

1 Student, CSE, Sphoorthy Engineering College, Telangana, India

2 Asst.Prof, CSE, Sphoorthy Engineering College, Telangana, India

3 Student, CSE, Sphoorthy Engineering College, Telangana, India

4 Student, CSE, Sphoorthy Engineering College, Telangana, India

5 Student, CSE, Sphoorthy Engineering College, Telangana, India

6 Student, CSE, Sphoorthy Engineering College, Telangana, India

## Abstract

Healthcare systems are increasingly adopting digital solutions to assist medical professionals in early disease diagnosis and patient monitoring. Existing symptom-based diagnostic systems predominantly focus on single-disease prediction, which fails to reflect real-world clinical scenarios where patients frequently present with overlapping symptoms and comorbid conditions. This paper proposes a Smart E-Healthcare System that leverages multi-label machine learning techniques to predict multiple diseases simultaneously from patient-reported symptoms. The system employs a fuzzy symptom-matching algorithm based on Levenshtein distance similarity to handle input variations and typographical errors, combined with a SHAP-inspired explainability module that provides per-symptom contribution scores to enhance clinical transparency and trust. The prediction engine supports eight major diseases—Diabetes, Hypertension, Asthma, Thyroid Disorder, Anemia, Migraine, Arthritis, and Common Cold—each characterized by ten clinically relevant symptoms with associated importance weights. Model performance is evaluated using appropriate multi-label metrics including Hamming Loss, Precision, Recall, and F1-score. The backend is implemented using Node.js with Express.js and a MySQL relational database, while the frontend delivers a responsive, mobile-friendly interface with integrated medicine recommendations and location-based doctor discovery. Experimental evaluation demonstrates a symptom recognition accuracy of 91%, Hamming Loss of 0.08, and macro-averaged F1-score of 0.87, with prediction response times under 50 milliseconds. The system addresses key challenges in e-health including security, scalability, and usability, making it a viable tool for preliminary multi-disease diagnostic support in real-world clinical environments.

**Keyword:** e-healthcare, multi-disease prediction, multi-label classification, clinical decision support, fuzzy matching, Levenshtein distance, symptom analysis, explainable AI, Hamming Loss, F1-score

## 1. INTRODUCTION

Healthcare systems are increasingly adopting digital solutions to assist medical professionals in early disease diagnosis and patient monitoring. With the rapid growth of medical data, Machine Learning (ML) techniques have shown great potential in supporting clinical decision-making [1]. However, most existing healthcare diagnostic systems focus on predicting a single disease at a time, which does not reflect real-world medical scenarios where patients often exhibit multiple symptoms and coexisting conditions [2]. Recent advancements in multi-label classification have enabled the prediction of multiple diseases

simultaneously, improving diagnostic accuracy and efficiency [3]. Despite these advancements, many existing systems still suffer from limited accuracy, poor interpretability, and lack of trust among healthcare professionals [4]. Moreover, traditional models fail to consider symptom interdependencies and often ignore appropriate evaluation metrics for multi-label prediction, such as Hamming Loss and F1-score, which are essential for assessing the true performance of multi-label classifiers [5].

Therefore, this paper proposes a Smart E-Healthcare System that leverages multi-label machine learning techniques to predict multiple diseases from patient symptoms. The proposed system aims to enhance diagnostic accuracy, improve decision support for doctors, and provide scalable healthcare solutions suitable for real-world clinical environments. The key contributions of this work are:

A multi-label disease prediction engine using fuzzy symptom matching with Levenshtein distance, supporting simultaneous prediction of eight major diseases.

- A SHAP-inspired explainability module providing per-symptom contribution scores to improve clinical transparency and trust.
- Evaluation using appropriate multi-label metrics including Hamming Loss, Precision, Recall, and macro-averaged F1-score.
- A secure, full-stack web application with user authentication, medicine recommendations, and location-based doctor discovery.
- A persistent diagnosis history module enabling longitudinal patient monitoring.

The remainder of this paper is organized as follows. Section II reviews related work and identifies the research gap. Section III defines the problem statement, objectives, and scope. Section IV describes the system architecture. Section V details the prediction algorithm. Section VI covers implementation details. Section VII presents evaluation results. Section VIII discusses security considerations. Section X concludes the paper.

## 2. Related work

### A. Symptom-Based Single-Disease Prediction

Several researchers have explored the application of machine learning in healthcare diagnostics. Patel et al. [6] developed a symptom-based disease prediction system using traditional classification models including Decision Trees, Naïve Bayes, and Support Vector Machines. While their system achieved reasonable accuracy for individual diseases, their approach was strictly limited to single-disease prediction and could not handle patients presenting with overlapping or comorbid conditions. Similarly, Garg et al. [11] evaluated the effects of computerized CDSS on practitioner performance and found that single-label systems frequently misclassify patients with multiple concurrent conditions reducing overall diagnostic reliability.

### B. Deep Learning Approaches

Verma et al. [7] applied deep learning models, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for automated disease diagnosis from clinical text and structured symptom data. Although their models demonstrated high accuracy on benchmark datasets, they did not consider multi-label scenarios or symptom overlap between diseases. Furthermore, deep learning models typically require large annotated datasets and significant computational resources, limiting their applicability in lightweight web-based deployments.

### C. Multi-Label Classification in Healthcare

Recent studies by Kumar and Singh [8] introduced multi-label classification techniques in healthcare, demonstrating improved performance in predicting multiple diseases simultaneously. Their work applied Binary Relevance and Label Powerset methods to symptom datasets and showed that multi-label approaches outperform single-label classifiers in comorbidity scenarios. However, these systems lacked proper evaluation using multi-label-specific metrics such as Hamming Loss and macro-averaged F1-score, making it difficult to assess their

true predictive performance. Additionally, most existing approaches do not provide sufficient interpretability, which reduces trust among healthcare professionals [10].

#### D. Explainable AI in Healthcare

The importance of explainability in medical AI has been highlighted by numerous studies [4], [10]. SHAP (SHapley Additive explanations) values provide a unified framework for feature attribution in ML models [4]. Ribeiro et al. [10] proposed LIME (Local Interpretable Model-agnostic Explanations) as an alternative approach. Both methods have been applied to clinical decision support to improve physician trust. Our system implements a lightweight SHAP-inspired mechanism that computes per-symptom contribution scores in real time without the computational overhead of full SHAP computation.

#### E. Identified Research Gap

From the above studies, it is observed that existing healthcare diagnostic system either focus on single disease prediction or lack proper multi-label evaluation and interpretability. Specifically: (1) single-label systems cannot handle comorbid conditions; (2) multi-label systems that do exist fail to use appropriate evaluation metrics such as Hamming Loss and F1-score; and (3) most systems provide no explainability mechanism, reducing clinical trust. This creates a clear gap for a reliable, interpretable, multi-disease prediction system that is evaluated using appropriate multi-label metrics and deployed as an accessible web-based platform.

### 3. PROBLEM DEFINITION

#### A. Problem Statement

Existing healthcare diagnostic systems fail to accurately predict multiple diseases simultaneously from patient symptoms, leading to reduced diagnostic efficiency and limited clinical trust. Single-label classifiers cannot model the reality of comorbid conditions, and the few multi-label systems that exist lack proper evaluation using multi-label metrics and do not provide interpretable outputs. There is a need for a robust smart healthcare system that supports multi-disease prediction using effective machine learning techniques, appropriate evaluation metrics, and explainable outputs suitable for clinical decision support.

#### B. Objectives

*The specific objectives of this work are:*

- 1) To design a smart e-healthcare system for multi-disease prediction from patient-reported symptoms.
- 2) To apply multi-label machine learning algorithms, specifically fuzzy symptom matching with weighted confidence scoring, for symptom-based diagnosis.
- 3) To evaluate model performance using appropriate multi-label metrics including Hamming Loss, Precision, Recall, and macro-averaged F1-score.
- 4) To improve accuracy and reliability of diagnostic decision support through SHAP-inspired symptom contribution explainability.
- 5) To develop a secure, scalable, and user-friendly web application integrating diagnosis, medicine recommendations, and doctor discovery.

#### C. Scope of the Project

The scope of the proposed system is defined as follows:

- Focuses on symptom-based disease prediction using patient-reported textual symptom input.
- Supports simultaneous prediction of multiple diseases (multi-label output) across eight disease categories.
- Intended for use as a clinical decision support tool, not as a replacement for professional medical diagnosis.
- Scalable for real-world healthcare environments through a RESTful web architecture and relational database back-end.

- Evaluation is conducted on controlled test cases; formal clinical trials are identified as future work.

#### 4. SYSTEM ARCHITECTURE

##### A. Overview

The Smart E-Health System follows a three-tier architecture comprising a presentation layer (frontend), an application layer (backend server), and a data layer (MySQL database).

- **Presentation Layer** : HTML5/CSS3/JavaScript views for login, dashboard, medicines, and doctors.
- **Application layer** : Node.js v 14+ with Express.js v4.18, providing RESTful API endpoints and session management.
- **Data Layer**: MYSQL 5.7+ relational database with users and prediction tables.

##### B. Frontend Modules

- 1) **Authentication Module (login.html, auth-script.js)**: User registration and login with client-side and server-side validation.
- 2) **Dashboard Module(dashboard.html, dashboard-script.js)**: Core diagnostic interface embedding the HealthcareAI multi-label prediction engine.
- 3) **3)Medicines Module (medicines.html, medicines-script.js)**: Searchable medicine catalog with dosage information across 8 disease categories.
- 4) **4) Doctors Module (doctors.html, doctors-script.js)**: Location-based doctor discovery with specialty filtering and direct contact links.

##### C. Symptom Processing

The raw user input string is tokenized by splitting on commas and whitespace, followed by lowercasing and trimming. Each token is matched against the full symptom vocabulary using Equation 1, producing a normalized recognized symptom set  $S_u = \{s_1, s_2, \dots, s_k\}$ .

##### D. Multi-Label Confidence Scoring

For each disease  $d$  with symptom set  $S_d$  and importance weights  $\{w_i\}$ , the confidence score is:

$$C(d) = \frac{\sum_{s_i \in S_u \cap S_d} w_i}{\sum_{s_j \in S_d} w_j} \times 100 \quad (2)$$

Diseases with  $C(d) < 20\%$  are filtered out. All remaining diseases are returned as the multi-label prediction output, ranked in descending order of  $C(d)$ . This threshold-based multi-label approach allows simultaneous prediction of multiple comorbid conditions.

##### E. Symptom Contribution (Explainability)

For each predicted disease  $d$ , a SHAP-inspired per-symptom contribution score is computed:  $\phi_i(d) = \frac{w_i}{\sum_{s_j \in S_u \cap S_d} w_j} \times 100 \quad (3)$

where  $\phi_i(d)$  is the percentage contribution of matched symptom  $s_i$  to the prediction of disease  $d$ . These scores are displayed alongside each prediction to provide interpretable, clinically meaningful explanations.

#### F. Algorithm Summary

##### Algorithm 1 Multi-Label Disease Prediction

Require: Symptom string  $I$ , knowledge base  $K$ , threshold  $\tau = 0.6$ , confidence cutoff  $\delta = 20$

Ensure: Ranked multi-label predictions with confidence and contributions

- 1:  $S_u \leftarrow \text{tokenize}(I)$
- 2:  $S_u \leftarrow \text{fuzzyMatch}(S_u, K, \tau)$
- 3:  $\text{results} \leftarrow []$
- 4: for each disease  $d \in K$  do
- 5:  $C(d) \leftarrow \text{computeConfidence}(S_u, d)$
- 6: if  $C(d) \geq \delta$  then

```

7: (d) ← computeContributions(Su,d)
8: results.append({d, C(d), Φ(d)})
9: end if
10: end for
11: return sort(results, by C(d) ↓

```

## 5. PREDICTION ALGORITHM

### A. Disease Knowledge Base

The system maintains a structured knowledge base of eight diseases, each defined by ten clinically, relevant symptoms and associated importance weights  $w_i \in [0,1]$ . The diseases are: Diabetes, Hypertension, Asthma, Thyroid Disorder, Anemia, Migrane, Arthritis, and common cold. Each disease has 10 symptoms, yielding a total vocabulary of 80 symptom entries across the knowledge base.

### B. Levenshtein Similarity

To handle typographical errors and spelling variations in user-entered symptoms, the system computes a normalized Levenshtein similarity between user input token  $u$  and each known symptom  $s$ :

$$\text{sim}(u, s) = 1 - \frac{\text{dL}(u, s)}{\max(|u|, |s|)} \quad (1)$$

where  $\text{dL}(u, s)$  is the Levenshtein edit distance and  $|u|, |s|$  denote string lengths. A similarity threshold  $\tau=0.6$  is applied; tokens with  $\text{sim}(u, s) \geq \tau$  are accepted as a match.

## 6. IMPLEMENTATION

### A. Technology Stack

- **Runtime:** Node.js v14+
- **Web Framework:** Express.js v4.18.2
- **Database:** MySQL 5.7+ via mysql2 v3.6.0
- **Authentication:** bcryptjs v2.4.3 (10 salt rounds), express-session v1.17.3
- **Frontend:** HTML5, CSS3, Vanilla JavaScript (ES6+)
- **Additional:** cors v2.8.5, body-parser v1.20.2

### B. Authentication Flow

User registration involves server-side validation, uniqueness checks on username and email, bcrypt hashing with 10 salt rounds, and insertion into the users table. Login involves credential retrieval, bcrypt comparison, and session creation. All protected routes verify req. session. userId before serving content.

### C. Prediction Engine Deployment

The HealthcareAI class is deployed on the client side (browser JavaScript), enabling sub-50ms predictions without server round-trips. Prediction results are persisted to the database via POST /api/prediction, enabling longitudinal diagnosis history tracking per user.

### D. Medicine Recommendation

The medicines module provides a static database of 40+ medicines organized by disease category. Each entry includes medicine name, recommended dosage, and drug type (e.g., oral tablet, inhaler, supplement). The interface supports full text search and disease-based filtering.

#### *E. Doctor Discovery*

The doctors module contains 25+ physicians across five cities. Each record includes name, specialty, phone, email, address, and estimated distance. Users can filter by city or specialty and access direct contact links.

#### *F. Responsive Design*

The UI employs CSS Grid and Flexbox with responsive breakpoints for desktop, tablet, and mobile layouts. A consistent gradient color scheme (#667eea to #764ba2) is applied across all pages.

## 7. EVALUATION AND RESULTS

### *A. Multi-Label Evaluation Metrics*

Evaluating multi-label classifiers requires metrics that account for partial label matches and label-set cardinality.

### *B. Prediction Performance*

The system was evaluated on 400 test cases (50 per disease), including correct spellings, common misspellings, and partial symptom inputs. Results are summarized. The Hamming Loss of 0.08 indicates that only 8% of label predictions are incorrect on average, demonstrating strong multi-label classification performance. The macro F1-score of 0.87 confirms balance precision and recall across all eight disease categories.

### *C. Usability Testing*

Informal usability testing with 10 participants (5 medical students, 5 general users) indicated that 90% successfully completed the full diagnosis workflow on first attempt, 80% rated the symptom contributions scores as “helpful” or “very helpful”, and average task completion time was under 3 minutes.

## 8. SECURITY ANALYSIS

Security is critical in healthcare applications due to the sensitivity of patient data. The system implements the following measures:

- 1) Password Hashing:** All passwords are hashed using bcryptjs with 10 salt rounds, making brute-force attacks computationally infeasible.
- 2) Session Management:** Express-session maintains server-side sessions with a 1-hour expiry. Sessions are destroyed on logout.
- 3) SQL Injection Prevention:** All database queries use parameterized prepared statements via the mysql2 driver.

**4) Input Validation:** Server-side validation enforces field completeness, email format, password confirmation, and uniqueness constraints.

**5) Protected Routes:** All authenticated endpoints verify req. session. userId before processing requests.

**6) CORS Configuration:** Cross-origin resource sharing is configured to restrict API access to trusted origins.

For production deployment, additional hardening is recommended: environment variable management for credentials, Redis-based session storage, HTTPS/TLS enforcement, CSRF token protection, and rate limiting on authentication end points.

## 9. LIMITATIONS AND FUTUREWORK

The current system has several limitations that motivate future research:

- **Knowledge base size:** The system supports 8 diseases. Expansion using ICD-10 codes and larger clinical datasets would increase coverage.
- **Trained ML model:** Replacing the rule-based scoring with a trained multi-label classifier (e.g., Random Forest with Binary Relevance, XGBoost with Label Powerset) on clinical datasets such as the UCI ML Repository or MIMIC-III would further improve accuracy.
- **Clinical validation:** Formal clinical trials comparing system predictions against physician diagnoses are needed to establish medical validity.
- **Real doctor data:** Integration with healthcare provider APIs would replace the static doctor dataset with live data.
- **Appointment booking:** Adding a scheduling module would complete the patient journey from diagnosis to consultation.
- **Mobile application:** A React Native or Flutter app would extend accessibility to smartphone users.
- **Multi-language support:** Localization would broaden reach to non-English-speaking populations.

## 10. CONCLUSION

This paper presented the Smart E-Healthcare System, a full-stack web application that leverages multi-label machine learning techniques to predict multiple diseases simultaneously from patient-reported symptoms. The system addresses a clear gap in existing literature: most diagnostic systems are limited to single-disease prediction, lack proper multi label evaluation metrics, and provide no explainability. The proposed system employs Levenshtein distance-based fuzzy symptom matching, a SHAP-inspired explainability module, and is evaluated using appropriate multi-label metrics including Hamming Loss (0.08) and macro-averaged F1-score (0.87). The Node.js/MySQL backend provides a secure, scalable foundation with comprehensive authentication and session management. The integrated medicine recommendation and location-based doctor discovery modules make the system a comprehensive clinical decision support tool. The system represents a practical, interpretable, and accessible solution for preliminary multi-disease diagnostic support in real-world healthcare environments.

## REFERENCES

- [1] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nature Medicine*, vol. 28, no. 1, pp. 31–38, Jan. 2022.
- [2] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333–359, 2011.
- [3] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.
- [4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.
- [5] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1–13, 2007.
- [6] R. Patel, A. Sharma, and N. Gupta, "Symptom-based disease prediction using machine learning classifiers," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, 2022.
- [7] S. Verma and R. Kumar, "Deep learning-based automated disease diagnosis from clinical symptom data," *Applied Soft Computing*, vol. 112, p. 107784, 2021.
- [8] A. Kumar and P. Singh, "Multi-label classification for simultaneous disease prediction in smart healthcare systems," *Expert Systems with Applications*, vol. 213, p. 119022, 2023.
- [9] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *BMJ*, vol. 330, no. 7494, p. 765, Apr. 2005.
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [11] A. Garg et al., "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes," *JAMA*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [12] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.
- [13] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [14] World Health Organization, "World Health Statistics 2023: Monitoring Health for the SDGs," WHO Press, Geneva, 2023.